Amendments to Specification

Please replace paragraph [0040] with the following paragraph.

- - [0040] FIG. 17**A-B** is a schematic depicting the inverse PCR procedure for recovering genomic tags associated to vector or viral integration events. A method of cleaving said cellular DNA such that inserted DNA (with sequence known to the operator) is cleaved once and flanking cellular DNA of unknown sequence is cleaved again in the regions contiguous to the inserted piece of DNA. Cleavage of the DNA occurs in a fashion generating ends that permit the circularization of DNA fragments producing a molecule with the sequence known to the operator flanking both sides, and continuous with, a variable length of cellular DNA of unknown sequence. **(Fig. 17A)** The region containing the unknown DNA is then amplified and sequenced. **(Fig. 17B)**

Please amend paragraph [0094] with the following paragraph:

- - Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology-Information ~~**(http://www.hebi.nlm.nih.gov/).**~~ This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence

for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always>0) and N (penalty score for mismatching residues; always<0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).- -

Please amend paragraph [0137] at page with the following paragraph.

- - The preferred embodiment of the invention will use vectors (DNA, RNA, DNA/RNA hybrids etc.) that contain markers which may be sorted to include but not limited to cell surface displayed or cytoplasmic protein; lipid, lipoprotein, glycolipid, and glycoprotein targets that can be tagged with specific fluorescent, chemiluminescent, or bioluminescent compounds using labeled antibodies, direct chemical linkage and/or combination of direct and indirect tagging. These vectors (see FIG. 2A-K, 13A) use either the processes of illegitimate recombination, homologous recombination, and/or viral vectors to integrate said markers into the genomic DNA of target cells (the integrated vector serves as a molecular bar code). Alu sequences are

approximately 300 bp in length and are found on average every 3000 bp in the human genome. Alu or other highly repetitive sequences can be used to induce homologous recombination for insertion of the marker gene. The vectors will be delivered to the target cells via standard gene delivery methods to include but not limited to lipid mediated transfection (cationic, anionic, and neutral charged), activated dendrimers (~~**PolyFect.TM**~~. **POLYFECT™** Reagent, ~~**SuperFect.TM**~~. **SUPERFECT™** Reagent {Qiagen}), Phenylethyleneimide (PEI), receptor mediated transfection (fusogenic peptide/protein), calcium phosphate transfection, electroporation, particle bombardment, direct injection of naked-DNA, diethylaminoethyl (DEAE-dextran transfection) etc. Though the preferred embodiment is the use of plasmid based vectors, the use of other high efficiency viral vectors is not precluded. - -

Please replace paragraph [0177] with the following paragraph:

- - As shown in FIG. 17 **A-B**, recovery of genetic material from the cells to be analyzed, in this example cellular DNA (inclusive of, but not limited to, cellular DNA since complementary DNA derived from cellular RNA (cDNA) may be used), the composition of which is partially known to the operator by virtue of the inclusion of the sequences encoding the marker peptide. The genetic locus containing the inserted sequence (or producing the RNA containing inserted marker gene sequences) is known as the "tagged gene."